Clasificación de patrones

Práctica 4: Support vector machines

Nombre:

Grupo:

1. (sección 3.1) Incluye en la matriz adjunta los errores de train, test y validación tanto para el clasificador lineal como para el que usa el kernel gaussiano. ¿Qué valores de P y de *h* se han usado?

P=0.1 h=1

|  |  |  |  |
| --- | --- | --- | --- |
|  | Error de train | Error de test | Error de validación |
| SVM lineal | 0.0673913 | 0.0826087 | 0.0619565 |
| SVM kernel | 0.382971 | 0.383696 | 0.379348 |

1. (sección 3.1) Obtén los errores de train al variar los parámetros P y *h* al usar un kernel Gaussiano e inclúyelos en una gráfica bidimensional. Haz lo mismo con los errores de validación, y determina los valores óptimos de P y *h*.

DONE

1. (sección 3.1) Incluye el código generado para resolver el apartado anterior.

DONE

1. (sección 3.1) Usa el clasificador obtenido para los valores óptimos de *P* y *h* para determinar la probabilidad de error de test. Compare el resultado con el obtenido en el apartado 1 de esta memoria.
2. (sección 3.2) Incluye los 6 indicadores () sobre la base de datos de test y con el clasificador óptimo obtenido anteriormente.
3. (sección 3.2) Justifica para una BD compuesta por 400 vectores de tipo SPAM y 4600 vectores de tipo MAIL, por qué los cocientes  resultan más adecuados que  para medir la bondad del clasificador.
4. (sección 3.3) A partir de los resultados del punto anterior, ¿cuál es la matriz de confusión para la base de datos de test.?
5. (sección 3.3) Determina las probabilidades a priori:



1. (sección 3.3) Clasifica el vector aislado (V\_analisis) y comenta si se clasifica correctamente o incorrectamente. ¿Contenía el e-mail correspondiente la palabra "make"? ¿Y la palabra "address"?
2. (sección 3.3) Determina la fiabilidad de la decisión usando el teorema de Bayes. Para ello, si el vector aislado V\_analisis se clasifica como SPAM halle la probabilidad de que sea realmente SPAM:

   
o bien si se clasifica como MAIL, halle la probabilidad de que sea realmente MAIL:   
    
Utiliza la base de datos de test y los resultados proporcionados por el clasificador cómo cread conveniente, y explica de forma detallada el método utilizado.

1. (sección 3.3) Incluye el código generado para calcular la probabilidad en el punto anterior.